Improving Particulate Matter Source Apportionment for Health Studies: A Trained Receptor Modeling Approach with Sensitivity, Uncertainty and Spatial Analyses

Paige Tolbert, Jeremy Sarnat, Mitch Klein, James Mulholland, Amit Marmur and Armistead (Ted) Russell

Georgia Tech, Emory and the Georgia EPD

This research is funded by
U.S. EPA - Science To Achieve Results (STAR) Program
Grant # R833866

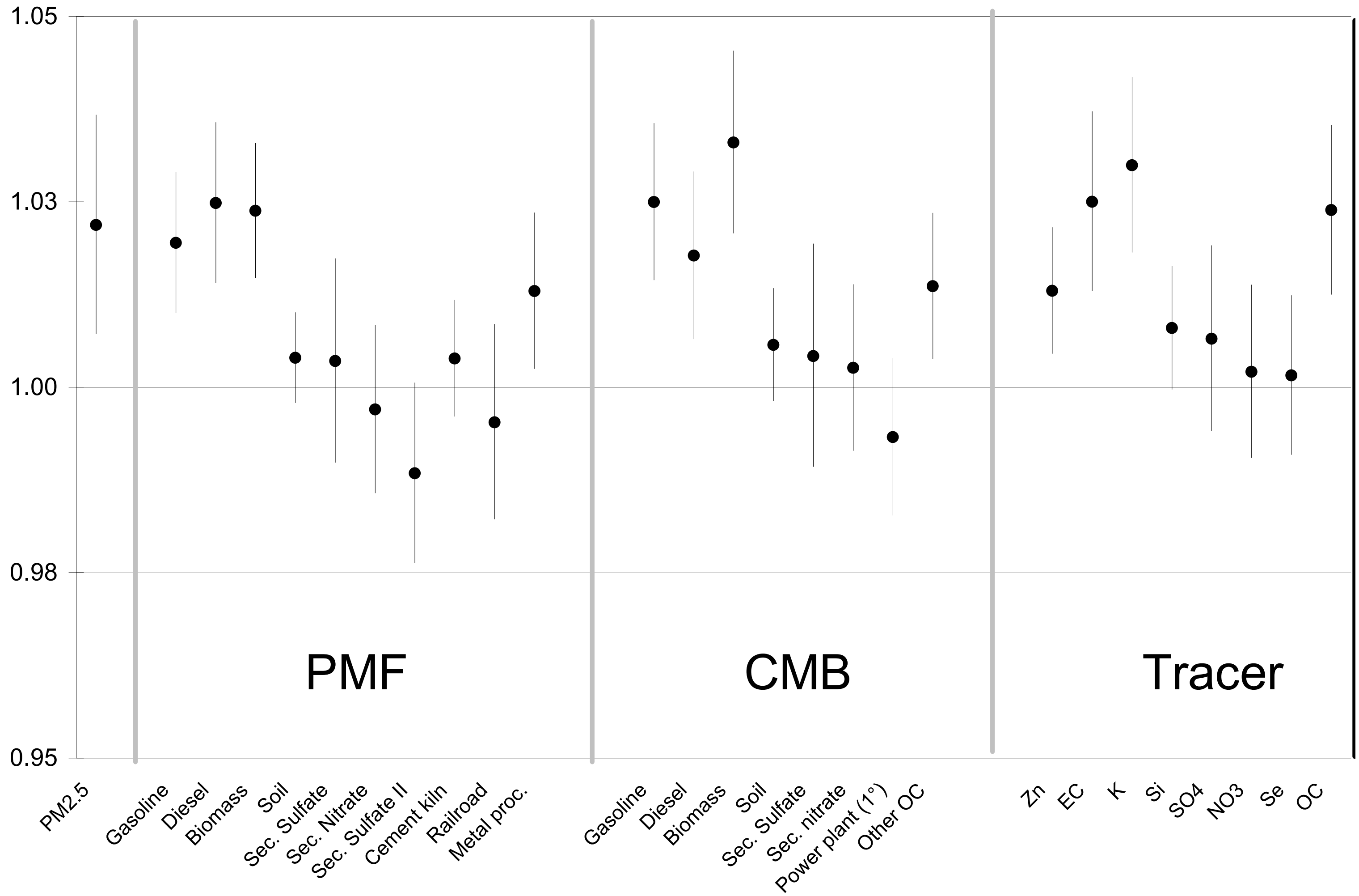
Acknowledgments

- US EPA
- Gyeongsangnam-do Province Government of Korea
- Southern Company

Air Quality Models and Health Impact Assessment

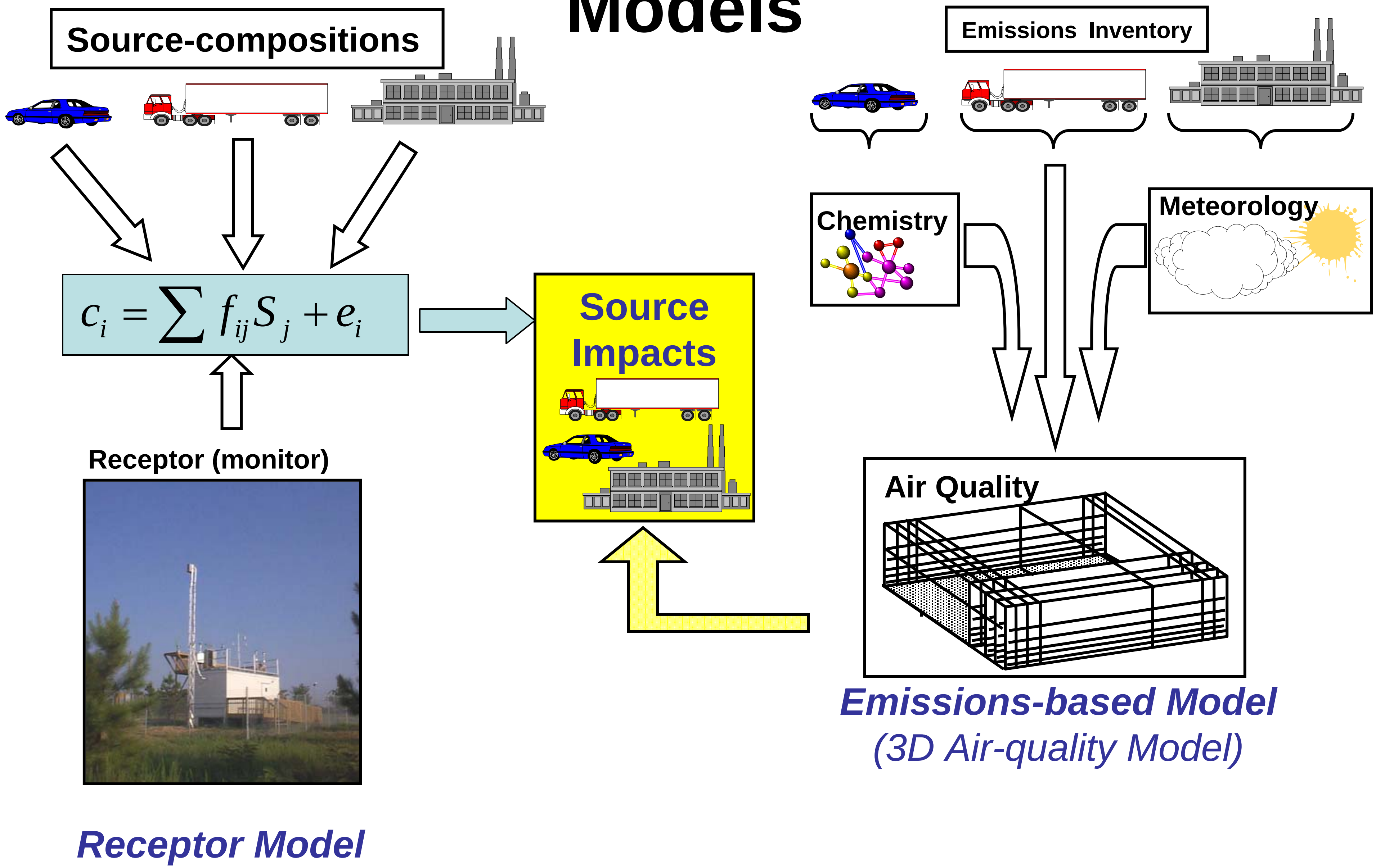
- (How) Can we use "air quality models" to help identify associations between ozone PM <u>sources</u> and health impacts?
  - Species vs. sources
  - Very different than for traditional air quality management
    - Though this is still a very important application

Use of Source Apportionment Results in Epidemiologic Studies



Sarnat, J. A.; et al.. *Environ. Health Perspect.* **2008,** *116*, (4), 459-466.

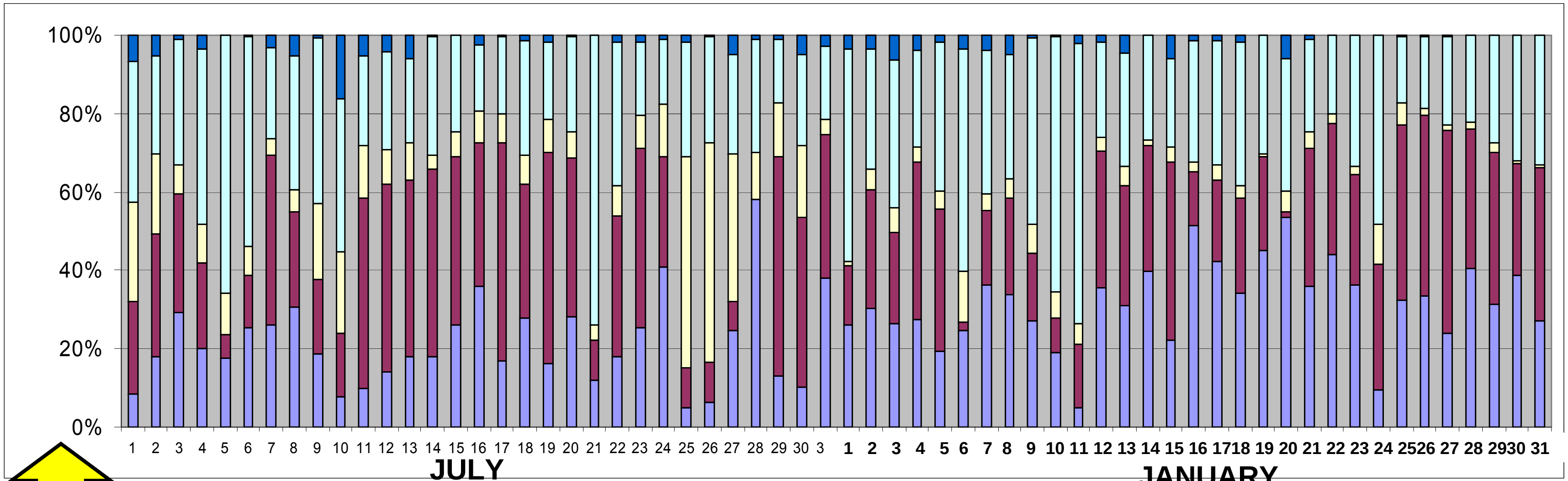
Receptor vs. Emissions-Based Models



The Problem!

Daily source apportionment (SA) results for Atlanta based on receptor and grid-based model results

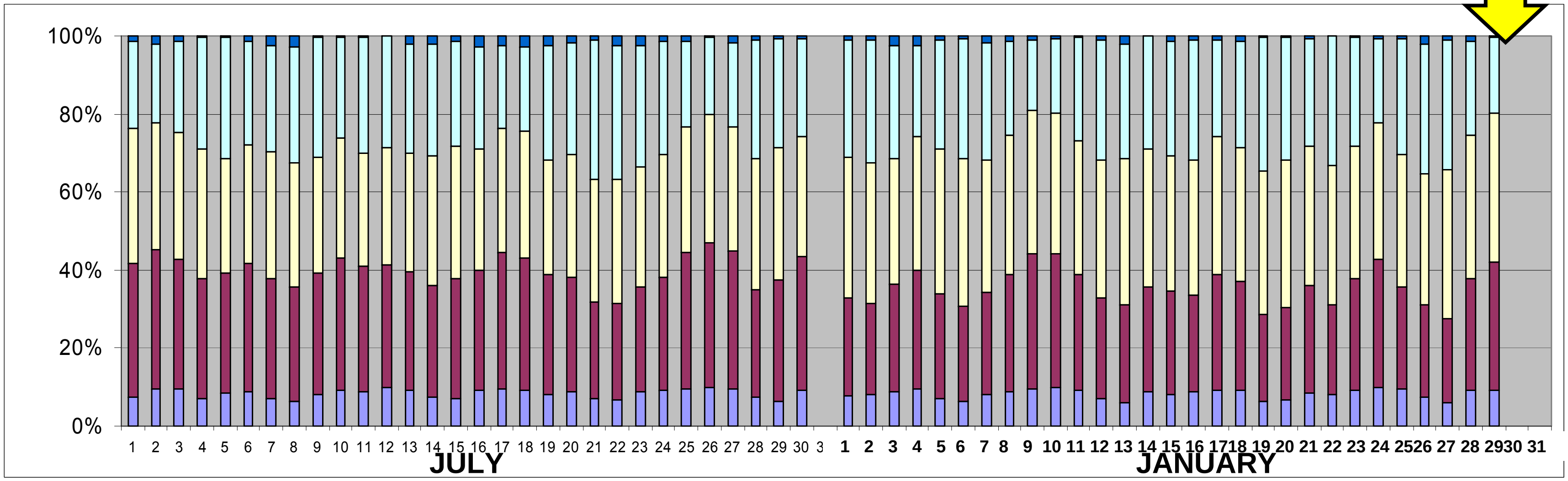
CMB



Significant daily variation

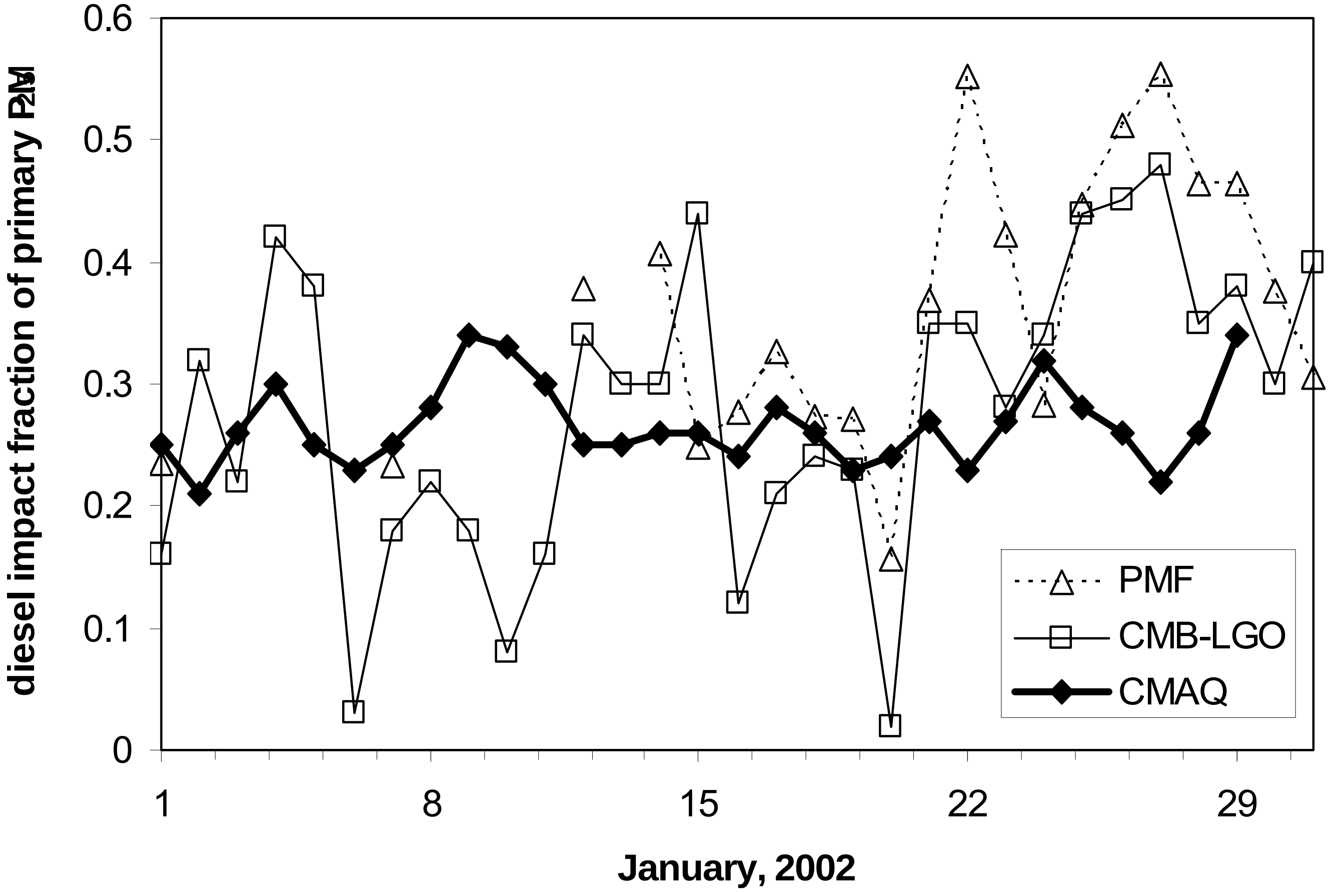
Little daily variation

CMAQ



■ LDGV ■ HDDV ■ SDUST ■ BURN ■ Coal

Diesel Impact Variation



Source correlations about 0.9 for CMAQ SA's

Both Results are Flawed

- Receptor
  - Too much day-to-day variability
    - Hard to imagine that diesel and coal burning impacts go to zero on some days and are significant on others
  - Missing sources
    - Little way around this
  - Source profiles uncertain and variable (plus that SOA issue)
- Grid-based
  - Too little variability
    - Tied to lack of small scale structure in met and emissions
    - Inconsistent with data
  - Inventories everywhere are uncertain (wrong)
- Can try to justify results
  - Our tests suggest arguments on both sides fail
- Use of source apportionment results for acute response epidemiologic analyses rely on getting day-to-day variability correct
  - Want to develop a more accurate SA for acute studies

Executive Summary

- Develop a flexible and extensible approach for source apportionment (SA)
  - Air quality management and epidemiologic studies
- Ensemble-trained approach
  - Integrate grid-based and multiple receptor modeling approaches
- Provide a tested method that directly addresses limitations in current SA methods, in particular variability, biases, and intensive resource requirements
  - Use SA results in epi studies of Atlanta and St. Louis

Limitations of Source Apportionment (SA) Approaches

- Receptor-based SA models:
  - biased estimates of primary source impacts
  - inability to identify or separate source impacts
  - excessive day-to-day variability
  - multiple zero impact days for sources that are known to be present (e.g. diesel vehicles, power plants)
  - results are representative for only the observation location
  - Some approaches resource intensive (detailed organic speciation)
- Emission-based chemical transport models (CTM):
  - large computational cost
  - results lack significant day-to-day variation in relative source impact
- Is it possible to improve results by taking an ensemble average of multiple approaches?
  - Then use ensemble results to train a receptor model

Ensemble SA and Training

- Develop SA results from weighted average of multiple methods over <u>limited period</u>
  - Chemical transport model (CTM)
  - Chemical mass balance models (CMB)
    - Regular (metals, ions, EC/OC)
    - Molecular Marker (MM: detailed organic speciation)
    - LGO (optimized profiles and constraints)
  - Positive matrix factorization (PMF)
  - Limited period allows using methods that are more resource intensive
  - Multiple methods allow estimating uncertainties
- Use ensemble results to develop optimized source profiles
  - Seasonally varying, location specific
- Use new profiles to calculate SA results over extended periods

Initial Application

1. **Ensemble source impacts** for July 2001 and January 2002 were developed by weighted averaging source impacts from a CTM (CMAQ) and multiple receptor-based approaches (CMB, CMB-MM, CMB-LGO, PMF).

2. **Ensemble-based source profiles (EBSPs)** for summer (July 2001) and winter (January 2002) were developed using ensemble-trained source impacts in CMB-LGO.

3. **New source impacts** were determined using CMB-LGO for a 12 month data set of daily PM2.5 measurements at the Atlanta, GA, Jefferson Street (JST) site using EBSPs.

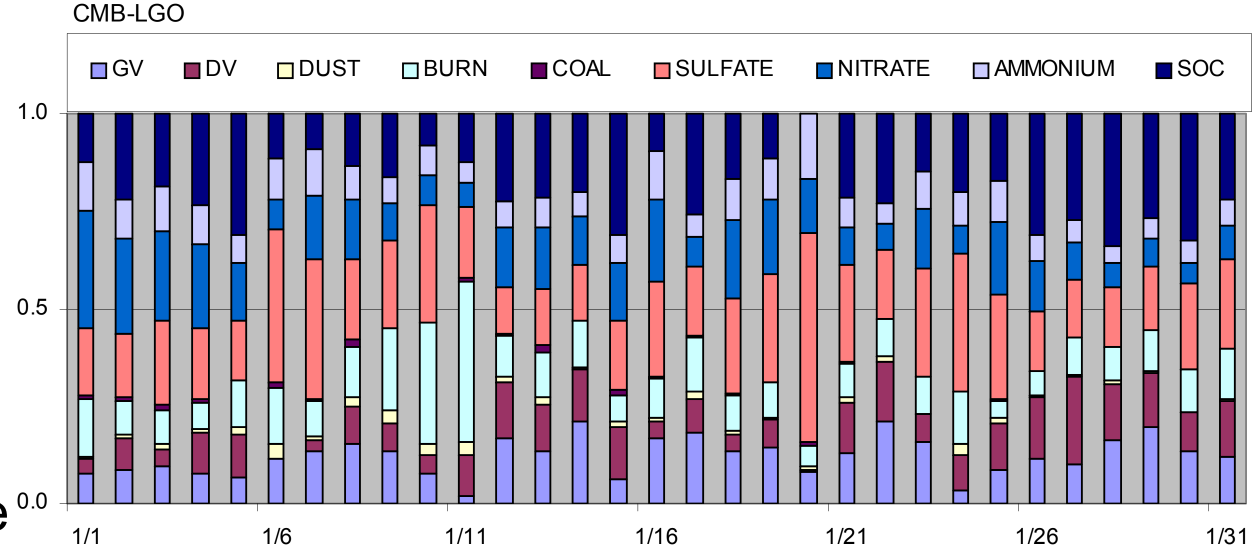
Step 1: Ensemble-Trained Source Impacts

- Run L individual SA methods (CMB, CMB-MM, CMB-LGO, PMF, CMAQ) to develop weighted source impacts

$$\bar{S}_j(t_k) = \frac{\sum_{l=1}^{L} w_{jl}(t_k) \bullet S_{lj}(t_k)}{\sum_{l=1}^{L} w_{jl}(t_k)} \qquad w_{jl} = \frac{1}{\sigma^2_{s_{lj}}}$$

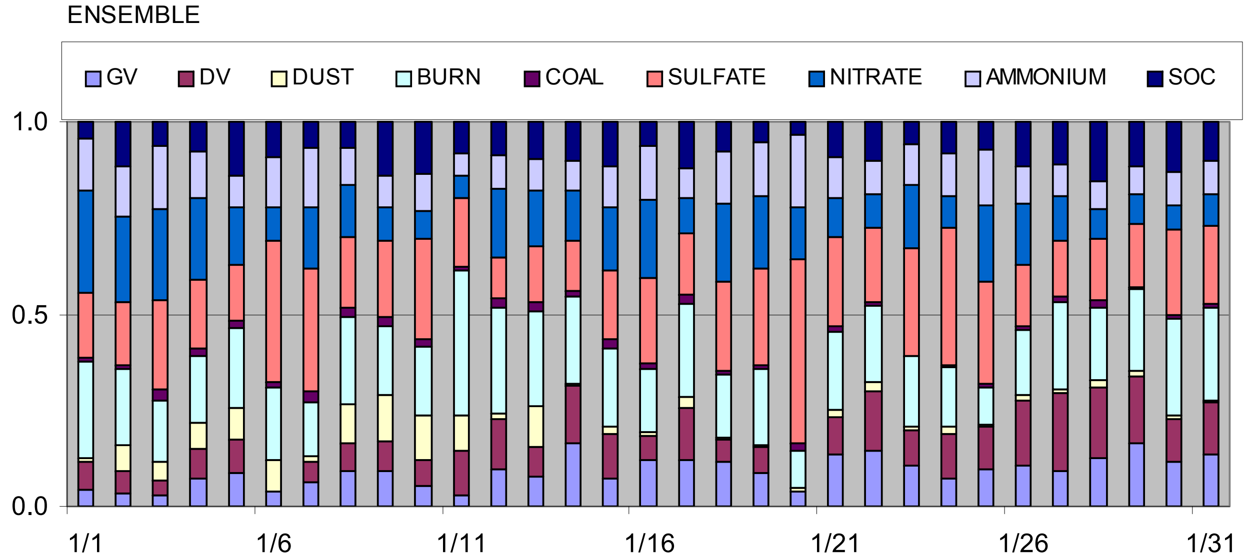
- $\bar{S}_j(t_k)$ is the ensemble-calculated impact of source j (in ug/m3) at time t_k
- $S_{lj}(t_k)$ is the impact developed by method *l*
- Weights, w_{jl}, are inversely proportional to uncertainty (derived from method application)

Ensemble results have less day-to-day variation in source impacts and fewer biases between observed and estimated PM2.5 mass compared to the original receptor model results.



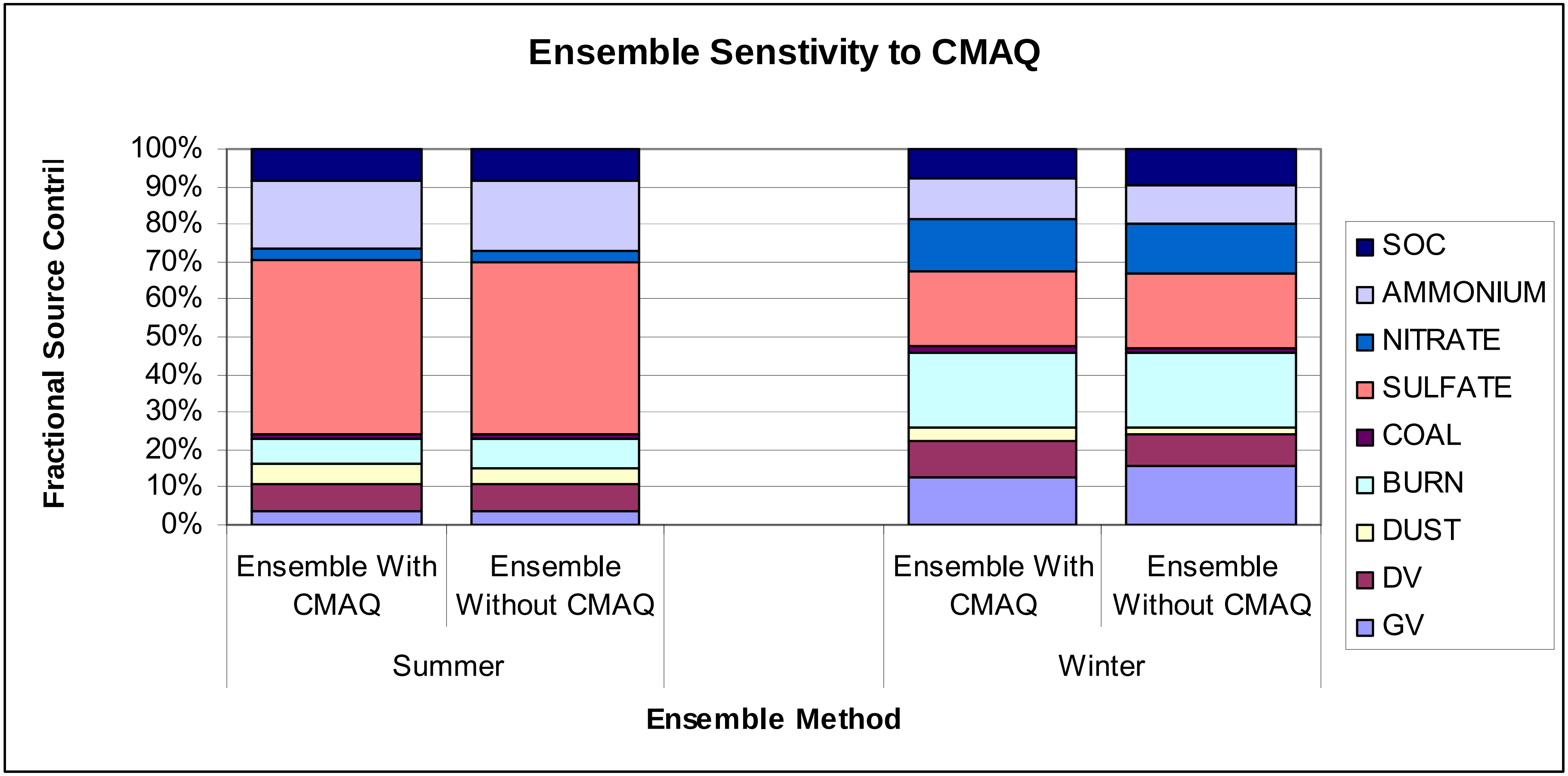
Daily fractional coutribution to PM2.5 in SA results developed from CMB-LGO method, at Jefferson St in Atlanta US, during January 2002

Ensemble results show increases in road dust, biomass burning, and coal combustion impacts, but SOC impacts decrease.



Daily fractional coutribution to PM2.5 in SA results developed from CMB-LGO method, at Jefferson St in Atlanta US, during January 2002

Ensemble Sensitivity to CMAQ



Step 2: Ensemble-Based Source Profiles (EBSPs)

- Problem: won't usually have multiple methods to ensemble for large data sets (e.g. 10 yrs).
  - CMAQ and CMB-MM
- Use ensemble source impacts from small data set (e.g. 1 month) to determine <u>ensemble based source profile</u>.
- EBSPs (f_{ij}) were treated as the unknown in the CMB equation and solved by minimizing least squares error.

$$C_{ik} = \sum_{j}^{J} \overline{f}_{ij} \cdot \overline{S}_{jk} + e_{ik}$$

$i = species$

$j = source$

$k = sample$

$$\chi^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{(C_{ik} - \sum_{j}^{J} \overline{f}_{ij} \cdot \overline{S}_{jk})^2}{\sigma_{c_{ik}}^2}$$

$\overline{f}_{ij} = \text{source profile}$

$\overline{S}_{jk} = \text{ensemble source impact}$

Step 3: New Source Impacts

- Develop new source impacts, S^*_{jk}, by minimizing least squares error using EBSPs
- CMB-LGO was run using EBSPs for a summer period (Mar - Oct) and a winter period (Jan - Feb, and Nov - Dec) and compared with measurement based source profiles (MBSPs)

$$C_{ik} = \sum_{j}^{J} \overline{f}_{ij} \cdot S^{*}_{jk} + e_{ik}$$

$$i = species$$

$$j = source$$

$$k = sample$$

$$\chi^2_k = \sum_{i=1}^{I} \frac{(C_{ik} - \sum_{j}^{J} \overline{f}_{ij} \cdot S^{*}_{jk})^2}{\sigma^2_{c_{ik}}}$$

$$\overline{f}_{ij} = EBSPs$$

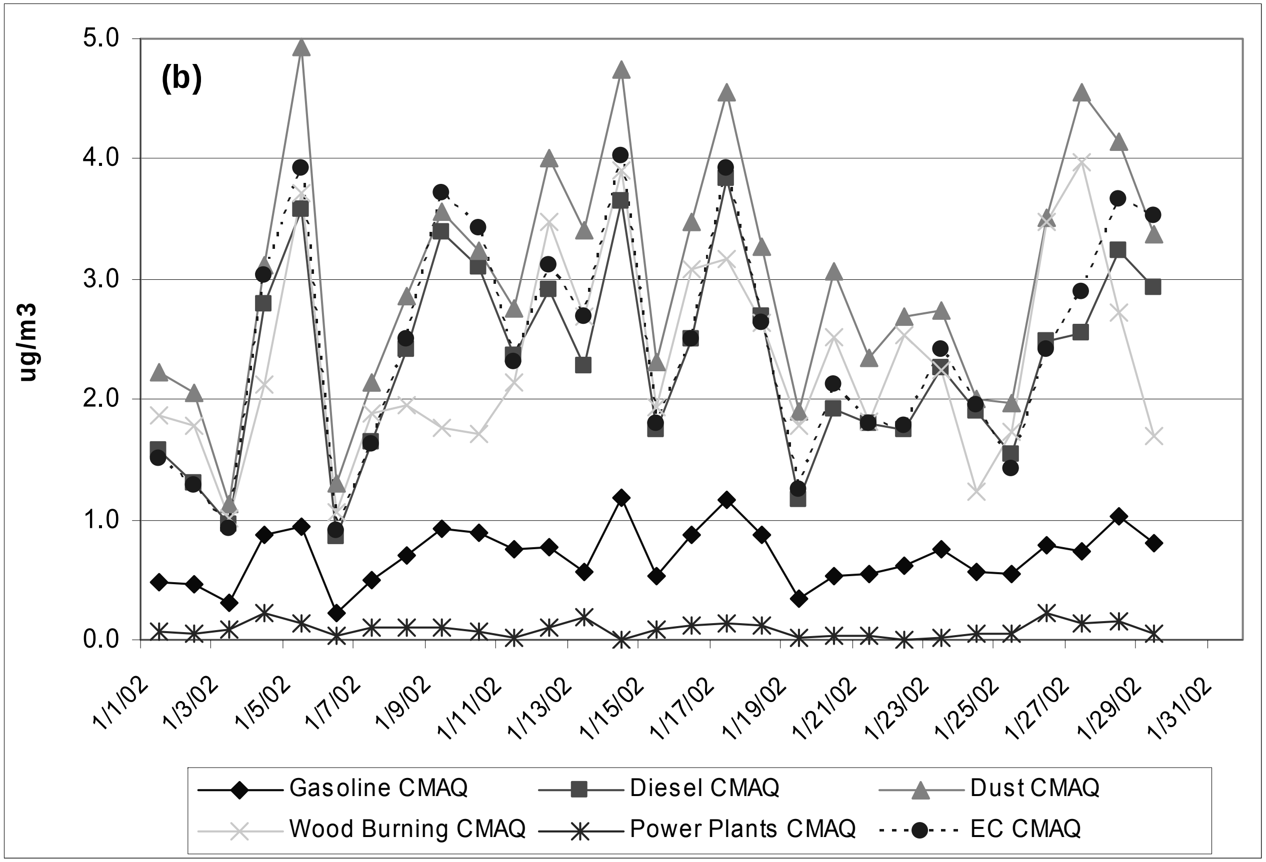
$$S^{*}_{jk} = \text{new source impacts}$$

Conclusions

- Ensemble-based source apportionment method developed to address limitations in current methods
  - Initial application to Atlanta
- Ensembling decreases variability and number of zero impact days
  - Assessed impact of including CMAQ results
- Ensemble-trained/based source profiles (EBSPs) developed for summer and winter
  - Results suggest seasonal variability in OC:EC ratios in profiles
- Application of EBSPs decreased variability, improved performance
  - Increased biomass burning and road dust impacts, decreased SOA in winter
- Future work will include
  - Applying the method to longer time periods and other locations
  - Assessing variability and refining the ensemble method
  - Using different approaches to estimating weights and assessing uncertainties
  - Conduct spatial analyses
  - Apply to more routine monitoring data
  - Incorporating source impacts into epidemiology studies

CMAQ Results

Source impacts at Jefferson Street, SEARCH site, Atlanta



Correlations between sources high (R ~ 0.9)